AD-A099 310 MARYLAND UNIV COLLEGE PARK F/G 12/1
INCOMPLETE AND FALSE IDENTIFICATION DISTRIBUTIONS: GROUP SCREEN--ETC(U)
MAY 81 S KOTZ, N L JOHNSON N00014-81-K-0301
UNCLASSIFIED NL

1 of 1

END
DATE FILMED
6-81
DTIC

(12)

28 May 81

Contract belongs to Maryland Univ. per Karen Bearington (ONR) 696-4508

60W TID

INCOMPLETE AND FALSE IDENTIFICATION DISTRIBUTIONS: GROUP SCREENING MODELS

by

Samuel Kotz
University of Maryland, College Park

and

Norman L. Johnson
University of North Carolina at Chapel Hill

Abstract

In Johnson *et al.* (*Commun. Statist. Theor. Meth. A9(9)*, 917-922) and Johnson and Kotz (*Proc. ONR/ARO Reliability Workshop*, April 1981), the authors previously derived the distribution of the number of items observed to be defective in samples from a finite population, when false identification of defectives as well as incomplete identification is taken into account. The corresponding distributions of waiting times until a specified number of defective items is observed were also obtained. In the present paper, we extend some of these results to the case of "group screening" sampling schemes.

Key Words and Phrases: group screening; binomial distribution; compound distributions; faulty identification; hypergeometric distribution; sampling inspection; waiting time; incomplete identification.

1. Introduction and classification of faulty hypergeometric inspection models.

In recent papers (Johnson *et al.* (1980) and Johnson and Kotz (1981a)), the authors developed several models for incomplete and false identification distributions, originally motivated by applications in auditing (Sorkin (1977)) and in quality control. These models can be viewed as a new variant of the damage models introduced by Rao and Rubin (1964), which have been extensively studied in the literature. (See Johnson and Kotz (1981b) for a survey of damage models and their relation to faulty inspection models.)

For completeness and readers' convenience, we shall briefly describe the main results offered in Johnson *et al.* (1980) and Johnson and Kotz (1981a).

1a) Incomplete identification.

Consider a sample of size n without replacement from a lot of size N conforming X defective (or nonconforming) items, when inspection detects such items with probability p ($0 < p \le 1$). It is assumed that no "correct" items are classified as defectives. In this model, the overall distribution of the total number of *identified* defectives, Z say, is found to be a compound binomial distribution.

$\text{Binomial}(Y,p) \underset{Y}{\wedge} \text{Hypergeometric}(n,X,N)$, where $\wedge$ denotes the compounding operation (Johnson and Kotz (1969, p. 184)) and Y denotes the *actual* (unobservable) number of defective items in a random sample (without replacement) of size n.

The formula for the s^{th} descending factorial moment of Z is

$$\mu_{(s)} = E(Z^{(s)}) = n^{(s)} X^{(s)} p^s / N^{(s)} \tag{1}$$

where $c^{(s)} = c(c-1)\ldots(c-s+1)$.

In particular,

$$E(Z) = pnX/N \ , \tag{2}$$

which is the mean of a hypergeometric distribution with parameters (n,Xp,N), formally representing the distribution of defectives in a sample (without replacement) of size n from a population of size N containing pX defectives.

The variance V(Z) can be written as

$$\mathrm{Var}(Z) = p^2 n \, \frac{N-n}{N-1} \, \frac{X}{N} (1 - \frac{X}{N}) + p(1-p) \frac{NX}{N} = p^2 \, \mathrm{Var}(Z \mid p=1) + p(1-p) \frac{nX}{N} \ , \tag{3.1}$$

or alternatively in the two following forms:

$$\mathrm{Var}(Z) = \frac{n(N-n)}{(N-1)} \, \frac{pX}{N} (1 - \frac{pX}{N}) + \frac{(n-1)}{(N-1)} \, p(1-p) \frac{nX}{N} \tag{3.2}$$

or

$$\mathrm{Var}(Z) = \frac{npX}{N} (1 - \frac{pX}{N}) - \frac{n(n-1)}{N(N-1)} \, p^2 X (1 - \frac{X}{N}) \ . \tag{3.3}$$

These show that V(Z) is not less than the variance of the hypergeometric with parameters (n,pX,N), but cannot exceed the variance of a binomial with parameters $(n, \frac{pX}{N})$.

The corresponding waiting time distribution of the number M of drawings (with replacement) of items needed to produce a ($\le$ X) defective items *recognised as such* (properized as $P(M=m)/P(M \le N)$) is the compound distribution

$$\text{Negative Hypergeometric}(Y,X,N) \underset{Y}{\wedge} \text{Truncated } (Y \le X) \text{ Negative Binomial}(a, p^{-1}-1) \ .$$

(The negative binomial is truncated from above at Y = X because there are no more than X defective items.)

It seems to be difficult to obtain exact expressions for the moments of M. However, if the truncation to values $Y \leq X$ is neglected, the s^{th} *ascending* factorial moment of M is given by

$$E(M^{[s]}) \doteq p^{-s} a^{[s]} (N+1)^{[s]} / (X+1)^{[s]} , \tag{4}$$

where $M^{[s]} = M(M+1)\ldots(M+s-1)$. In particular,

$$E[M] \doteq \frac{a(N+1)}{p(X+1)} \tag{5.1}$$

and

$$\mathrm{Var}(M) \doteq \frac{a(N+1)}{p^2(X+1)^2\ (X+2)} [(N+2)(X+1) - p(X+1)(X+2) - a(N-X)] . \tag{5.2}$$

(See Johnson *et al.* (1980) for more details.)

1b) <u>False and incomplete identification</u>.

In Johnson and Kotz (1981a), the model described in (a) was extended by allowing for a probability, p', of erroneously deciding that an item is defective when really it is not. (In the purely incomplete model, $p' = 0$.)

In this case, the overall distribution of the total number of items *called "defectives"*, Z, is the compound distribution

$$\mathrm{Binomial}(Y,p) + \mathrm{Binomial}(n-Y,\ p') \underset{Y}{\wedge} \mathrm{Hypergeometric}(n,X,N)$$

(the two binomial variables are mutually independent).

The r^{th} descending factorial moment of Z is in this case given by

$$\mu_{(r)}(Z) = \frac{n^{(r)}}{N^{(r)}} \sum_{j=0}^{r} \binom{r}{j} p^j p'^{\,r-j} X^{(j)} (N-X)^{(r-j)} ; \tag{6}$$

in particular,

$$E(Z) = n\bar{p}/N$$

where $\bar{p} = \{Xp + (N-X)p'\}/N$, and the variance is

$$\mathrm{Var}(Z) = \frac{n\bar{p}}{N}(1 - \frac{\bar{p}}{N}) - \frac{n(n-1)}{(N-1)} \frac{X}{N} (1 - \frac{X}{N})(p-p')^2$$

(c.f. corresponding expression for the variance of Z in the case 1a). Tables of the distribution of Z for $p = .75(.05).95$; $p' = 0(.025).1$; $N = 100$, $X = 5, 10, 20$; $N = 200$, $X = 10, 20, 40$; and $n = 10$ are presented in Johnson and Kotz (1981a). More detailed tables may be obtained by writing to S. Kotz. The distributions are quite sensitive to the values of p', but not to the values of the ratio $\frac{X}{N}$. In fact, as N and X are increased proportionately to each other with $X/N = \lambda$, say, the other parameters (n,p,p') remaining constant, the distribution of Z tends to a binomial with parameters $n, XN^{-1}p + (1 - XN^{-1})p'$. The waiting time distribution (i.e. the distribution of the number of items M, say, needed to be inspected one at a time until a predetermined number a of items have been assessed as "defective") seems to be difficult to derive. Using a conditioning argument, Johnson and Kotz (1981a) obtained close approximations and bounds on the values of E(M) and Var(M) in this case.

These are

$$E(M) \doteq a\bar{p}^{-1}[1 + \frac{1}{N}[1 + \frac{2}{\bar{p}} - \frac{1}{N\bar{p}^2}\{Xp^2 + (N-X)p'^2\}] \qquad (7.1)$$

and

$$\frac{a(1-\bar{p})}{\bar{p}^2}[1 + \frac{1}{N} - \frac{3+a}{N\bar{p}}] \le \mathrm{Var}(M) \le \frac{a(1-\bar{p})}{\bar{p}^2}(1 + \frac{1}{N}) \, , \qquad (7.2)$$

As $N \to \infty$, E(M) approaches $a\bar{p}^{-1}$ and the variance tends to $a\bar{p}^{-2}(1-\bar{p})$. These are the mean and variance, respectively, of the (negative binomial) waiting time distribution for occurrence of a "successes" in independent trials with probability of success equal to $\bar{p}$ at each trial.

These results can easily be generalized to the case of stratified populations where the lot is divided into k strata of sizes $X_1, X_2, \ldots, X_k$ ($\sum X_j = N$) such that for any chosen individual in the j^{th} stratum, the probability of "detection as defective" (whether this is really so or not) is p_j. The case considered above corresponds to $k = 2$, $p_1 = p$, and $p_2 = p'$. See Johnson and Kotz (1981a) for more details.

2. Group screening model involving incomplete and false identification.

Further interesting distributions arise in connection with "group screening" (Dorfman (1943)), in which groups of units can be tested for the existence of one or more defective units among them. This can be practicable, for example, when testing liquids for presence of contaminants, and is then suggested as a possible way of reducing the average total amount of testing.

Suppose that material from n units is mixed and tested for presence of "defective" material. If a negative result ("no defectives") is obtained, no further action is taken, but if there is a positive result, each unit is tested separately.

Let p_0, p_0' denote the probabilities of obtaining correct or incorrect positive results, respectively, at the first test. As before, p,p' denote the probabilities of correct or incorrect positive results, respectively, when units are tested individually; X,N denote the number of defective units and the total number of units in the population respectively, and Y denotes the actual number of defective units among the n tested.

The overall probability of obtaining a positive result on the first test is

$$\{1 - P(Y=0)\}p_0 + P(Y=0)p_0' = \{1 - \frac{(N-X)^{(n)}}{N^{(n)}}\}p_0 + \frac{(N-X)^{(n)}}{N^{(n)}} p_0' . \quad (8)$$

As before, Z will denote the number of units called "defective" as a result of the test.

When $Y = 0$, the conditional distribution of Z is binomial with parameters n, p' plus "added zeroes" (corresponding to a negative result on the first test):

$$\begin{aligned} P(Z=0|Y=0) &= 1 - p_0' + p_0'(1-p')^n \\ P(Z=z|Y=0) &= p_0'\binom{n}{z}p'^z(1-p')^{n-z} \qquad (z = 1,2,\ldots,n) . \end{aligned} \quad (9)$$

When $Y = y > 0$, the conditional distribution of Z is that of the sum of two independent binomial variables with parameters $(y,p),(n-y, p')$ plus "added zeroes":

$$\begin{aligned} P(Z=0|Y=y) &= 1 - p_0 + p_0(1-p)^y (1-p')^{n-y} \qquad (y > 0) \\ P(Z=z|Y=y) &= p_0 \sum_{j=0}^{z} \binom{y}{j}p^j(1-p)^{y-j} \binom{n-y}{z-j}p'^{z-j}(1-p')^{n-y-z+j} \\ &\qquad (y > 0;\ z = 1,2,\ldots,n) . \end{aligned} \quad (10)$$

The overall distribution of Z is obtained by compounding (9) and (10) with a hypergeometric distribution (parameters n,X,N) for Y. The r^{th} factorial moment of Z is

$$E[Z^{(r)}] = n^{(r)}\left\{\frac{p_0}{N^{(r)}} \sum_{i=0}^{r} \binom{r}{i}X^{(i)}(N-X)^{(r-i)}p^i p'^{r-i} - \frac{(p_0-p_0')p'^r(N-X)^{(n)}}{N^{(n)}}\right\} . \quad (11)$$

Formula (11) can be obtained by noting that formally the distribution of Z is a mixture of

(a) Binomial(Y,p) + Binomial(n-Y, p') $\underset{Y}{\wedge}$ Hypergeometric(n,X,N) with probability p_0,

(b) Binomial(n,p') with "probability" $(p_0'-p_0)P(Y=0)$, and

(c) 0 with probability $(1-p_0)P(Y>0)+(1-p_0')P(Y=0)$.

(Note that the "probability" for (b) can be negative; indeed, it is quite likely that $p_0' < p$.)

In particular,

$$E[Z] = n(p_0\bar{p} - Pp') , \tag{12.1}$$

where as before, $\bar{p} = XN^{-1}p - (1-XN^{-1})p'$; $P = (p_0-p_0')(N-X)^{(n)}/N^{(n)}$, and

$$\begin{aligned} Var(Z) = n(n-1)[&\frac{p_0}{N-1}\{N\bar{p}^2 - N^{-1}(Xp^2+\overline{N-X}\cdot p'^2)\} - Pp'^2] \\ &+ n(p_0\bar{p} - Pp') - n^2(p_0\bar{p} - Pp')^2 . \end{aligned} \tag{12.2}$$

In general, it would seem that $p_0 > p_0'$ just as $p > p'$, since we would expect (hope) that the probability of correct decision would exceed that of incorrect decision. It may well happen that $p_0 < p$ since detection of a defective may be more difficult with the mixture of material from separate units. More complicated distributions will be obtained if it is supposed that p_0 depends on the value of Y (the number of defective units). It does not seem unreasonable to suppose that p_0 might increase with Y.

Acknowledgement

Samuel Kotz's work was supported by the U.S. Office of Naval Research under Contract N00014-81-K-0301. Norman L. Johnson's work was supported by the National Science Foundation under Grant MCS-8021704.

References

Dorfman, R. (1943). The detection of defective members of large populations. *Ann. Math. Statist. 14*, 436-440.

Johnson, N.L. and Kotz, S. (1969). *Distribution in Statistics -- Discrete Distributions*. New York: John Wiley and Sons.

Johnson, N.L. and Kotz, S. (1981a). Faulty inspection distributions -- some generalizations. (To be published in *Proc. of ONR/ARO Reliability Workshop*, April 1981) *Institute of Statistics Mimeo Series #1335*, University of North Carolina at Chapel Hill.

Johnson, N.L. and Kotz, S. (1981b). Advances in discrete distributions (1969-1980). To appear in *ISI Review*.

Johnson, N.L., Kotz, S. and Sorkin, H.L. (1980). Faulty inspection distributions. *Commun. Statist. A9(9)*, 917-922.

Rao, C.R. and Rubin, H. (1964). On a characterization of the Poisson distribution. *Sankhyā A26*, 295-298.

Sorkin, H.L. (1977). An Empirical Study of Three Confirmation Techniques: Desirability of Expanding the Respondent's Decision Field. Ph.D. Thesis, University of Minnesota.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 6	2. GOVT ACCESSION NO. AD-A099310	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Incomplete and False Identification Distributions: Group Screening Models.		5. TYPE OF REPORT & PERIOD COVERED 9 TECHNICAL rept.
		6. PERFORMING ORG. REPORT NUMBER 15
7. AUTHOR(s) 10 Samuel Kotz [redacted] Norman L. Johnson		8. CONTRACT OR GRANT NUMBER(s) [redacted] N00014-81-K-0301, NSF [redacted]-MCS-8021704 New
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 11
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Office of Naval Research Statistics and Probability Program (Code 436) Arlington, VA 22217		12. REPORT DATE May [redacted]81
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12 11		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release -- Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

group screening; binomial distribution; compound distributions; faulty identification; hypergeometric distribution; sampling inspection; waiting time; incomplete identification.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In Johnson *et al.* (*Commun. Statist. Theor. Meth. A9(9)*, 917-922) and Johnson and Kotz (*Proc. ONR/ARO Reliability Workshop*, April 1981), the authors derived the distribution of the number of items observed to be defective in samples from a finite population, when false identification of defectives as well as incomplete identification is taken into account. The corresponding distributions of waiting times until a specified number of defective items is observed were also obtained. In the present paper, we extend some of these

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

219500 . JW

20. results to the case of "group screening" sampling schemes.

DATE
ILMED
-8